

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : <b>G06K 9/20, G06T 7/00</b>		<b>A1</b>	(11) International Publication Number: <b>WO 00/30023</b> (43) International Publication Date: 25 May 2000 (25.05.00)
(21) International Application Number: PCT/US99/27372 (22) International Filing Date: 17 November 1999 (17.11.99) (30) Priority Data: 09/195,361 17 November 1998 (17.11.98) US 09/264,921 9 March 1999 (09.03.99) US (71) Applicant: HOLOPLEX, INC. [US/US]; Suite 102, 600 South Lake Avenue, Pasadena, CA 91106 (US). (72) Inventors: PU, Allen; 1114 Las Riendas Way, Pasadena, CA 91107 (US). QIAO, Yong; 808 Wigwam Avenue, Arcadia, CA 91006 (US). ESCOBAR, Nelson; Apartment 302, 1310 E. Orange Grove Boulevard, Pasadena, CA 91104 (US). ICHIRIU, Michael; 472 S. El Molino Avenue #101, Pasadena, CA 91101 (US). (74) Agents: MEYER, Sheldon, R. et al.; Fliesler Dubb Meyer and Lovejoy LLP, Suite 400, Four Embarcadero Center, San Francisco, CA 94111-4156 (US).			(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: STEREO-VISION FOR GESTURE RECOGNITION			
(57) Abstract <p>A method and an apparatus to identify a gesture of a subject without the need of a fixed background. The apparatus includes a sensor and a computing engine. The sensor captures images of the subject. The computing engine analyzes the captured images to determine 3-D profiles of the subject, and the gestures of the subject. Information in the images not within a volume of interest is ignored in identifying the gesture of the subject.</p>			
<pre>graph TD     120((120)) --- 112((112))     112 --- 110((SUBJECT))     110 --- 112     112 --- 114((SOURCE))     112 --- 116((SENSOR))     116 --- 118((COMPUTING ENGINE))     114 --- 116     116 --- 118     118 --- 125[125]</pre>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

-1-

## STEREO-VISION FOR GESTURE RECOGNITION

The present invention relates generally to gesture recognition, and more specifically to using stereo-vision for gesture recognition.

5 To identify the gestures of a subject, typically, the subject's background should be removed. One way to remove the background is to erect a wall behind the subject. After images of the subject are captured, the fixed background--the wall--is removed from the images before the gestures are identified.

It should be apparent from the foregoing that the wall increases the cost of the setup and the complexity to identify the gestures of the subject.

-2-

SUMMARY OF THE INVENTION

The present invention identifies gestures of a subject without the need of a fixed background. One embodiment is through stereo-vision with a sensor capturing the images of the subject. Based on the images, and through ignoring  
5 information outside a volume of interest, a computing engine analyzes the images to construct 3-D profiles of the subject, and then identifies the gestures of the subject through the profiles. The volume of interest may be pre-defined, or may be defined by identifying a location related to the subject.

Only one sensor may be required. The sensor can capture the images  
10 through scanning, with the position of the sensor changed to capture each of the images. In analyzing the images, the positions of the sensor in capturing the images are taken into account. In another approach, the subject is illuminated by a source that generates a specific pattern. The images are then analyzed considering the amount of distortion in the pattern caused by the subject.

15 In one embodiment, the images are captured simultaneously by more than one sensor, with the position of at least one sensor relative to one other sensor being known.

In another embodiment, the subject has at least one foot, with the position of the at least one foot determined by a pressure-sensitive floor mat to help identify  
20 the subject's gesture.

The subject can be illuminated by infrared radiation, with the sensor being an infrared detector. The sensor can include a filter that passes the radiation.

In one embodiment, the volume of interest includes at least one region of interest, which, with the subject, includes a plurality of pixels. In analyzing the  
25 images to identify the gesture of the subject, the computing engine calculates the number of pixels of the subject overlapping the pixels of the at least one region of interest.

-3-

In another embodiment, the position and size of the at least one region of interest depend on a dimension of the subject, or a location of the subject.

In yet another embodiment, the present gesture of the subject depends on its prior gesture.

5        Other aspects and advantages of the present invention will become apparent from the following detailed description, which, when taken in conjunction with the accompanying drawings, illustrates by way of example the principles of the invention.

-4-

**BRIEF DESCRIPTION OF THE DRAWINGS**

FIG. 1 shows one embodiment illustrating a set of steps to implement the present invention.

FIG. 2 illustrates one embodiment of an apparatus of the present invention capturing an image of a subject.

FIG. 3 shows different embodiments of the present invention in capturing the images of the subject.

FIGS. 4A-C show one embodiment of the present invention based on the distortion of the specific pattern of a source.

FIG. 5 illustrates one embodiment of the present invention based on infrared radiation.

FIG. 6 shows different embodiments of the present invention in analyzing the captured images.

FIG. 7 shows one embodiment of a pressure-sensitive mat for the present invention.

FIG. 8 shows another embodiment of an apparatus to implement the present invention.

Same numerals in FIGS. 1-8 are assigned to similar elements in all the figures. Embodiments of the invention are discussed below with reference to FIGS. 1-8. However, those skilled in the art will readily appreciate that the detailed description given herein with respect to these figures is for explanatory purposes as the invention extends beyond these limited embodiments.

-5-

DETAILED DESCRIPTION

In one embodiment, the present invention isolates a subject from a background without depending on erecting a known background behind the subject. A three dimensional (3-D) profile of the subject is generated with the subject's gesture identified. The embodiment ignores information not within a volume of interest, where the subject probably is moving inside.

FIG. 1 shows one approach 100 of using an apparatus 125 shown in FIG. 2 to identify the gestures of the subject 110. At least one sensor 116, such as a video camera, captures (step 102) a number of images of the subject for a computing engine 118 to analyze (step 104) so as to identify gestures of the subject.

In one embodiment, the computing engine 118 does not take into consideration information in the images outside a volume of interest 112. For example, information in the images too far to the sides or too high can be ignored, which means that certain information is removed as a function of distance away from the sensor. Based on the volume of interest 112, the subject is isolated from its background. The gesture of the subject can be identified without the need of a fixed background.

FIG. 3 shows different embodiments of the present invention in capturing the images of the subject. One embodiment depends on using more than one sensor (step 150) to capture images simultaneously. In this embodiment, the position of at least one sensor relative to one other sensor is known. The position includes the orientation of the sensors. For example, one position is pointing at a certain direction, and another position is pointing at another direction.

Based on the images captured, the computing engine, 118, using standard stereo-vision algorithms, analyzes the captured images to isolate and to generate a 3-D profile of the subject. This can be done, for example, by comparing the

-6-

disparity between the more than one image captured simultaneously, and can be similar to the human visual system. The stereo-vision algorithm can compute 3-D information, such as the depth, or a distance away from a sensor, or a location related to the subject. That location can be the center of the subject. Information in the images too far to the sides or too high from the location can be ignored, which means that certain information is removed as a function of distance away from the sensors. In this way, the depth information can help to set the volume of interest, with information outside the volume not considered in subsequent computation. Based on the volume of interest, the subject can be isolated from its background.

In another embodiment, only one sensor is necessary. In one approach, the sensor captures more than one image at more than one position (step 152). For example, the sensor is a radar or a lidar, which measures returns. The radar can capture more than one image of the subject through scanning. This can be through rotating or moving the radar to capture an image of the subject at each position of the radar. In this embodiment, to generate the 3-D profile of the subject, the computing engine 118 takes into consideration the position of sensor when it captures an image. Before the subject has substantially changed his gesture, the sensor would have changed its position and captured another image. Based on the images, the 3-D profile of the subject is constructed. The construction process should be obvious to those skilled in the art. In one embodiment, the process is similar to those used in the synthetic aperture radar fields.

In another embodiment, the image captured to generate the profile of the subject depends on illuminating the subject by a source 114 that generates a specific pattern (step 154). For example, the light source can project lines or a grid of points. In analyzing the images, the computing engine considers the distortion of the pattern by the subject.



-7-

FIGS. 4A-C show one embodiment of the present invention depending on the distortion of the specific pattern of a source. FIG. 4A shows a light pattern of parallel lines, with the same spacing between lines, generated by a light source. As the distance from the light source increases, the spacing also increases. FIG. 4B shows a ball as an example of a 3-D object. FIG. 4C shows an example of the sensor measurement of the light pattern projected onto the ball. The distance of points on the ball from the sensor can be determined by the spacing of the projected lines around that point. A point in the vicinity of a smaller spacing is a point closer to the sensor.

10 In another embodiment, to enhance the ability in isolating the subject from the unknown background, the source 114 illuminates (step 160 in FIG. 5) the subject with infrared radiation, and the sensor 116 is an infrared sensor. The sensor may also include a filter that passes the radiation. For example, the 3dB bandwidth of the filter covers all of the frequencies of the source. With the infrared sensor, the effect of background noises, such as sunlight, is significantly diminished, increasing the signal-to-noise ratio.

FIG. 6 shows different embodiments of the present invention in analyzing the captured images. In one embodiment, the volume of interest 112 is predefined, 170. In other words, independent of the images captured, the computing engine 118 always ignores information in the captured images outside the same volume of interest to construct a 3-D profile of the subject.

After constructing the profiles of the subject, the computing engine 118 can determine the subject's gestures through a number of image-recognition techniques. In one embodiment, the subject's gestures can be determined by the distance between a certain part of the body and the sensors. For example, if the sensors are in front of the subject, a punch would be a gesture from the upper part of the body. That gesture is closer to the sensors than the position of the center of the body.

-8-

Similarly, a kick would be a gesture from the lower part of the body that is closer to the sensors.

In one embodiment, the volume of interest 112 includes more than one region of interest, 120. Each region of interest occupies a specific 3-D volume of space. In one approach, the computing engine 118 determines the gesture of the subject based on the regions of interest occupied by the 3-D profile of the subject. Each region of interest can be for designating a gesture. For example, one region can be located in front of the right-hand side of the subject's upper body. A part of the 3-D profile of the subject occupying that region implies the gesture of a right punch by the subject.

One embodiment to determine whether a region of interest has been occupied by the subject is based on pixels. The subject and the regions of interest can be represented by pixels distributed three dimensionally. The computing engine 118 determines whether a region is occupied by calculating the number of pixels of the subject overlapping the pixels of a region of interest. When a significant number of pixels of a region is overlapped, such as more than 20%, that region is occupied. Overlapping can be calculated by counting or by dot products. In another embodiment, the gesture of the subject is identified through edge detection, 173. The edges of the 3-D profile of the subject are tracked. When an edge of the subject falls onto a region of interest, that region of interest is occupied. Edge detection techniques should be obvious to those skilled in the art, and will not be further described in this application.

One embodiment uses information on at least one dimension of the subject, such as its height or size, to determine the position and size of at least one region of interest. For example, a child's arms and legs are typically shorter than an adult's. The regions of interest for punches and kicks should be smaller and closer to a child's body than to an adult's body. By scaling the regions of interest and by

-9-

setting the position of the regions of interest, based on, for example, the height of the subject, this embodiment is able to more accurately recognize the gestures of the subject.

5 This technique of modifying the regions of interest based on at least one dimension of the subject is not limited to three dimensional imaging. The technique can be applied, for example, to identify the gestures of a subject in two dimensional images. The idea is that after the 2-D profile of a subject is found from the captured images, the positions and sizes of two dimensional regions of interest can be modified based on, for example, the height of the profile.

10 Another embodiment sets the location of at least one region of interest based on tracking the position, such as the center, of the subject. This embodiment can more accurately identify the subject's gesture while the subject is moving. For example, when the computing engine has detected that the subject has moved to a forward position, the computing engine will move the region of interest for the kick  
15 gesture in the same direction. This, for example, reduces the possibility of identifying incorrectly a kick gesture when the body of the subject, rather than a foot or a leg, falls into the region of interest for the kick gesture. Identification of the movement of the subject can be through identifying the position of the center of the subject.

20 This technique of tracking the position of the subject to improve the accuracy in gesture recognition is also not limited to three dimensional imaging. The technique can be applied, for example, to identify the gestures of a subject in two dimensional images. The idea again is that after the 2-D profile of a subject is found from the captured images, the positions of two dimensional regions of interest  
25 can be modified based on, for example, the center of the profile. \*

In yet another embodiment, the computing engine takes into consideration a prior gesture of the subject to determine its present gesture. Remembering the

-10-

temporal characteristics of the gestures can improve the accuracy of gesture recognition. For example, a punch gesture may be detected when a certain part of the subject is determined to be located in the region of interest for a punch. However, if the subject kicks really high, the subject's leg might get into the region of interest for a punch. The computing engine may identify the gesture of a punch incorrectly. Such confusion may be alleviated if the computing engine also considers the temporal characteristics of gestures. For example, a gesture is identified as a punch only if the upper part of the subject extends into the region of interest for a punch. By tracking the prior position of body parts over a period of time, the computing engine enhances the accuracy of gesture recognition.

This technique of considering prior gestures to identify the current gesture of a subject again is not limited to 3-D imaging. For example, the technique is that after the 2-D profile of a subject is found from the captured images, the computing engine identifies the current gesture depending on the prior 2-D gesture of the subject.

FIG. 7 shows one embodiment of a pressure-sensitive floor mat 190 for the present invention. The floor mat further enhances the accuracy of identifying the subject's gesture based on the foot placement. In the above embodiments, the sensor 116 can identify the gestures. However, sometimes in situation, from the perspective of the sensor, when a certain part of the subject occludes another part of the subject, there might be false identification. For example, if there is only one sensing element, and the subject is standing directly in front of it, with one leg directly behind the other leg. Under this situation, it might be difficult for the computing engine to identify the gesture of the other leg stepping backwards. The pressure-sensitive mat 190 embedded in the floor of the embodiment 125 solves this potential problem.

In FIG. 7, the pressure sensitive mat is divided into nine areas, with a center

-11-

floor-region (Mat A) surrounded by eight peripheral floor-regions (Mat B) in four prime directions and the four diagonal directions. In this embodiment, the location of the foot does not have to be identified very precisely. When a floor-region is stepped on, a circuit is closed, providing an indication to the computing engine that a foot is in that region. In one embodiment, stepping on a specific floor-region can provide a signal to trigger a certain event.

In one embodiment, the volume of interest 112 is not predefined. The computing engine 118 analyzes the captured images to construct 3-D profiles of the subject and its environment. For example, the environment can include chairs and tables. Then, based on information regarding the characteristics of the profile of the subject, such as the subject should have an upright body with two arms and two legs, the computing engine identifies the subject from its environment. From the profile, the engine identifies a location related to the subject, such as the center of the subject. Based on the location, the computing engine defines the volume of interest, 172. Everything outside the volume is ignored in subsequent computation. For example, information regarding the chairs and tables will not be in subsequent computation.

FIG. 8 shows one embodiment 200 of an apparatus to implement the present invention. More than one infrared sensor 202 simultaneously capture images of the subject, illuminated by infrared sources 204. The infrared sensors have pre-installed infrared filters. After images have been captured, a computing engine 208 analyzes them to identify the subject's gestures. Different types of movements by the subject can be recognized, including body movement, such as jumping, crouching, leaning forward and backward; arm movements such as punching, climbing, and hand motions; and foot movements such as kicking, moving toward, and backward. Then, the gestures of the subject can be reproduced as the gestures of a video game figure shown on the screen of a monitor 206.

-12-

In one embodiment, the screen of the monitor 206 shown in FIG. 8 is 50 inches in diameter. Both sensors are of the same height from the ground and are four inches apart horizontally. In this embodiment, a pre-defined volume of interest is 4 feet wide, 7 feet long and 8 feet high, with the center of the volume being located at 3.5 feet away in front of the center of the sensors and 4 feet above the floor.

The present invention can be extended to identify the gestures of more than one subject. In one embodiment, there are two subjects, and they are spaced apart. Each has its own volume of interest, and the two volumes of interest do not intersect. The two subjects may play a game using an embodiment similar to the one shown in FIG. 8. As each subject moves, its gesture is recognized and reproduced as the gesture of a video game figure shown on the screen of the monitor 206. The two video game figures can interact in the game, controlled by the gestures of the subjects.

Techniques using, such as radar, lidar and cameras, have been described. Other techniques may be used to measure, such as depth information, which in turn can determine volume of interest. Such techniques include using an array of ultrasonic distance measurement devices, and an array of infrared LEDs or laser diodes and detectors.

Other embodiments of the invention will be apparent to those skilled in the art from a consideration of this specification or practice of the invention disclosed herein. It is intended that the specification and examples be considered as exemplary only, with the true scope and spirit of the invention being indicated by the following claims.

-13-

CLAIMS

1

2 We claim:

3

4 1. A method for obtaining information regarding a subject (110)  
5 without the need of a fixed background, the method comprising the steps of:  
6 capturing (102) images of the subject; and  
7 analyzing (104) the captured images without considering information in the  
8 images outside a volume of interest (112) for obtaining information regarding the  
9 subject.

10

11 2. A method as recited in claim 1 wherein:  
12 the method is for identifying at least one gesture of the subject (110);  
13 the step of analyzing is for identifying at least one gesture of the subject  
14 (110); and  
15 the images are captured simultaneously by more than one sensor (116), with  
16 the position of at least one sensor relative to one other sensor being known.

17

18 3. A method as claimed in any preceding claim wherein the volume of  
19 interest (112) is pre-defined.

20

21 4. A method as claimed in any preceding claims wherein:  
22 at least a part of the volume of interest (112) and at least a part of the  
23 subject (110) are represented by a plurality of pixels; and  
24 the step of analyzing includes the step of calculating the number of pixels of  
25 the subject (110) overlapping the pixels of the volume of interest (112) to obtain  
26 information regarding the subject (110).

27

-14-

1           5.     A method as recited in claims 1, 2 or 3, wherein the step of analyzing  
2 includes the steps of:  
3           identifying the profile of the subject (110) based on the images; and  
4           determining whether an edge of the profile of the subject (110) is within the  
5 volume of interest (112).

6  
7           6.     A method as recited in any preceding claims wherein the position of  
8 at least a part of the volume of interest (112) depends on at least one dimension of  
9 the subject (110).

10  
11           7.     A method as recited in any preceding claims wherein the size of at  
12 least a part of the volume of interest (112) is scaled based on at least one dimension  
13 of the subject (110).

14  
15           8.     A method as recited in any preceding claims wherein at least one  
16 position of the volume of interest (112) depends on one position of the subject  
17 (110).

18  
19           9.     An apparatus (125) for obtaining information regarding a subject  
20 (110) without the need of a fixed background, the apparatus (125) comprising:  
21           a sensor (116) configured to capture images of the subject; and  
22           a computing engine (118) configured to analyze the captured images  
23 without considering information in the images outside a volume of interest for  
24 obtaining information regarding the subject.

25  
26           10.    An apparatus (125) as recited in claim 9, wherein  
27 the apparatus is configured for identifying at least one gesture of the subject



-15-

1 (110);

2 the computing engine is configured to analyze the captured image for  
3 identifying at least one gesture of the subject (110); and

4 the apparatus further comprises at least one additional sensor to  
5 simultaneously capture images of the subject, with the position of at least one  
6 sensor relative to one other sensor being known.

7

8 11. An apparatus (125) as recited in claims 9 or 10 wherein the volume  
9 of interest is pre-defined.

10

11 12. An apparatus (125) as recited in claims 9, 10 or 11 wherein:  
12 at least a part of the volume of interest (112) and at least a part of the  
13 subject (110) are represented by a plurality of pixels; and  
14 the computing engine (125) is configured to calculate the number of pixels  
15 of the subject overlapping the pixels of the volume of interest (112) to obtain  
16 information regarding the subject (110).

17

18 13. An apparatus (125) as recited in claims 9, 10 or 11 wherein the  
19 computing engine (125) is configured to  
20 identify the profile of the subject (110) based on the images; and  
21 determine whether an edge of the profile of the subject (110) is within the  
22 volume of interest (112).

23

24 14. An apparatus (125) as recited in claims 9, 10, 11, 12 or 13 wherein  
25 the position of at least a part of the volume of interest (112) depends on at least one  
26 dimension of the subject (110).

27

-16-

1           15.    An apparatus (125) as recited in claims 9, 10, 11, 12, 13 or 14  
2    wherein the size of at least a part of the volume of interest (112) is scaled based on  
3    at least one dimension of the subject (110).

4

5           16.    An apparatus (125) as recited in claims 9, 10, 11, 12, 13, 14 or 15  
6    wherein at least one position of the volume of interest (112) depends on one  
7    position of the subject (110).

$\sim 100$ 

Capture the images  $\sim 10^2$



analyze the images  
without considering  $\sim 10^4$   
information outside a  
volume of interest

Fig. 1

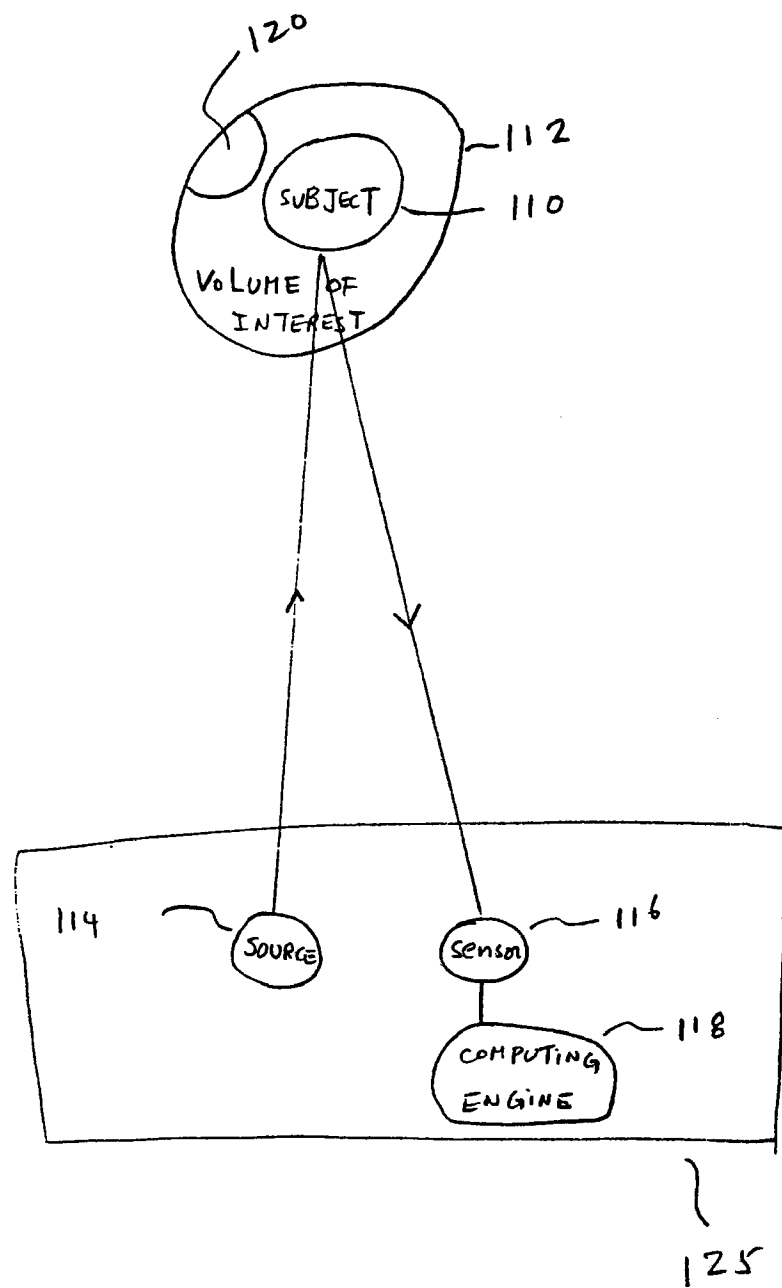


Fig. 2

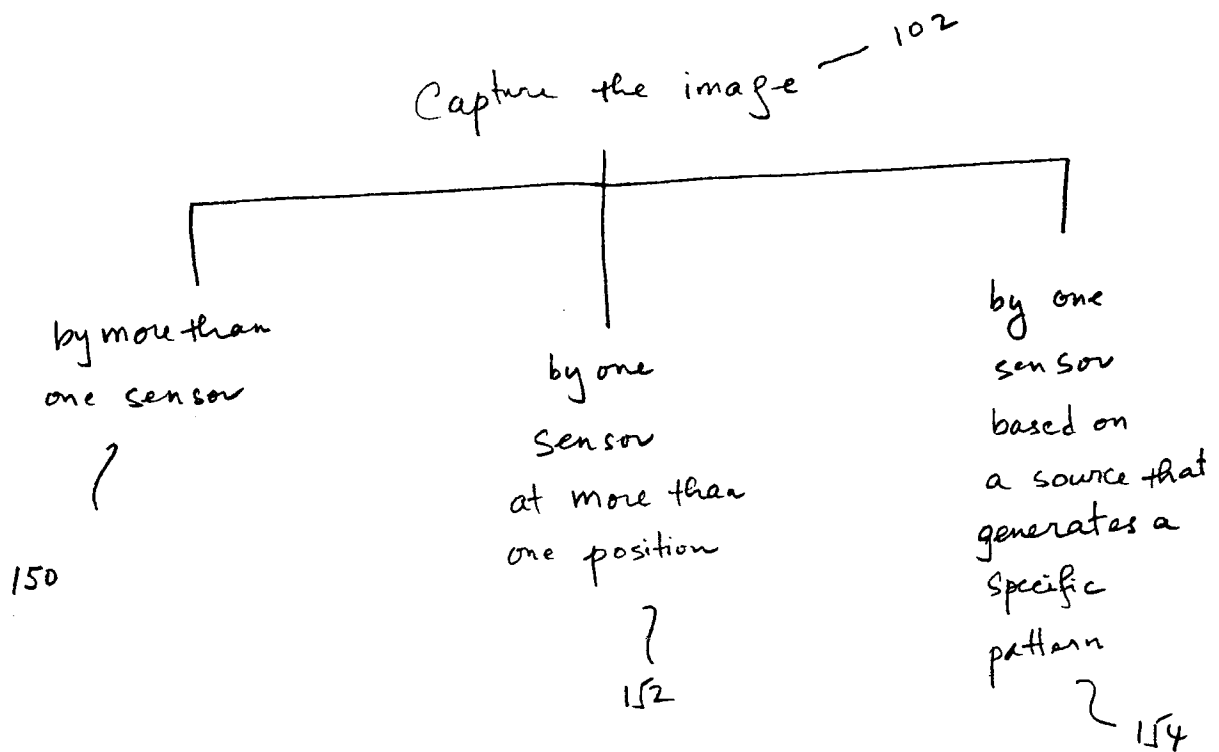


Fig. 3

Fig. 4A

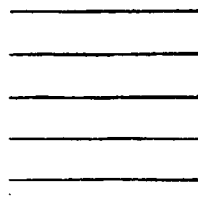


Fig. 4B

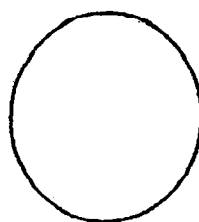
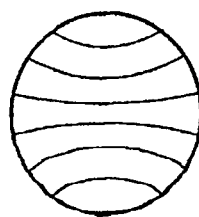


Fig. 4C



subject illuminated  $\sim 16^\circ$   
by infrared  
source

↓  
capture the  $\sim 16^\circ$   
infrared radiation

FIG. 5'

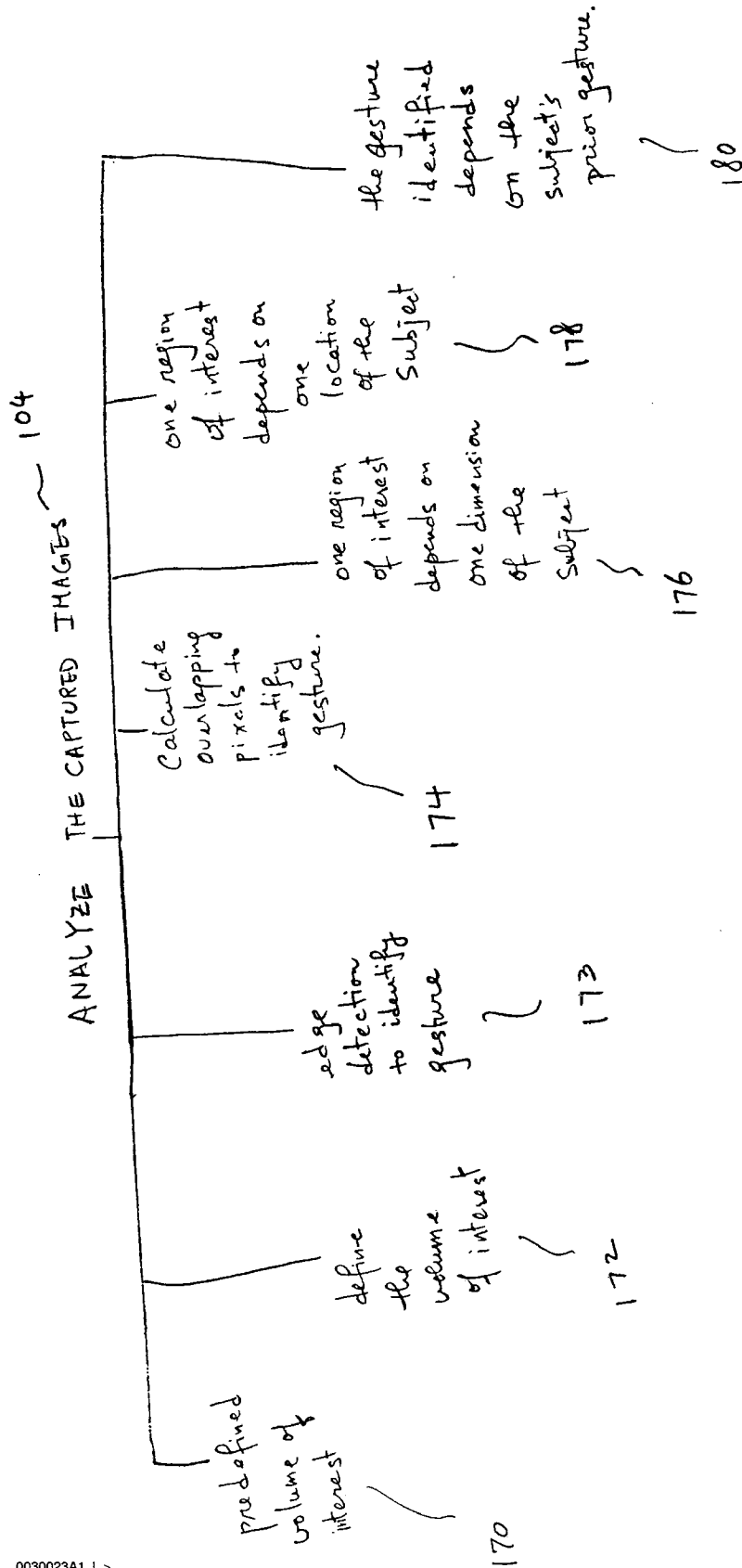


FIG. 6



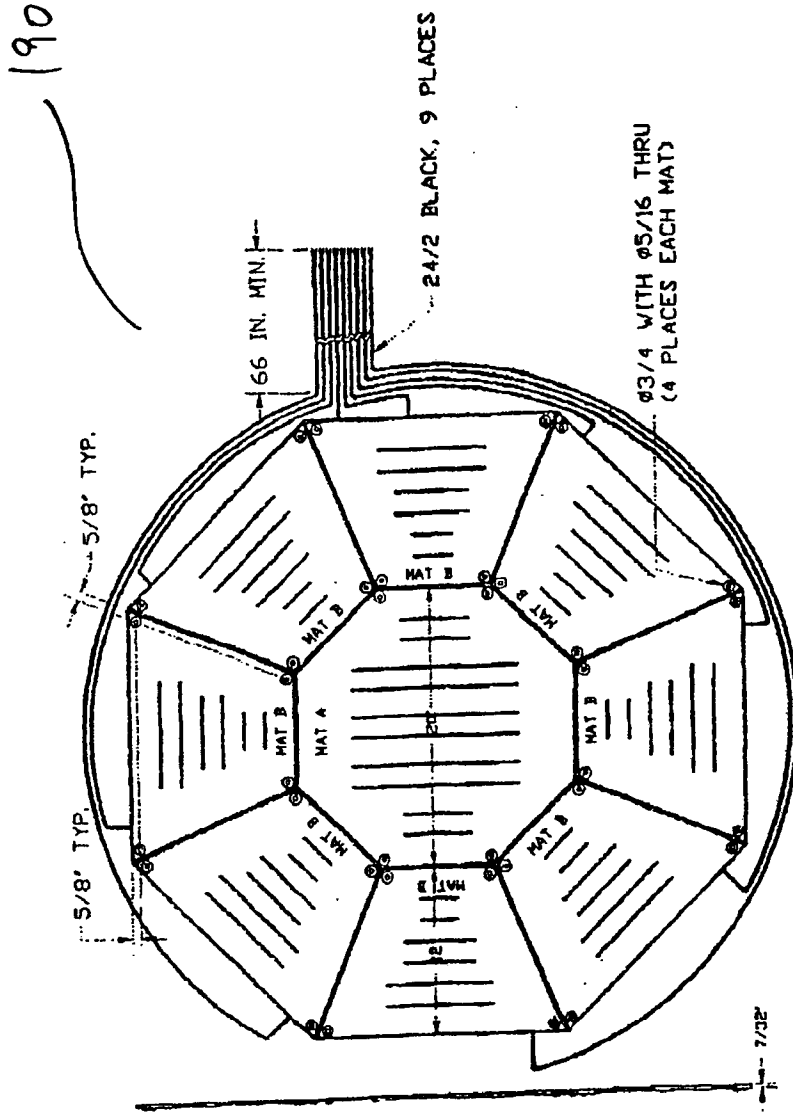


Fig. 7

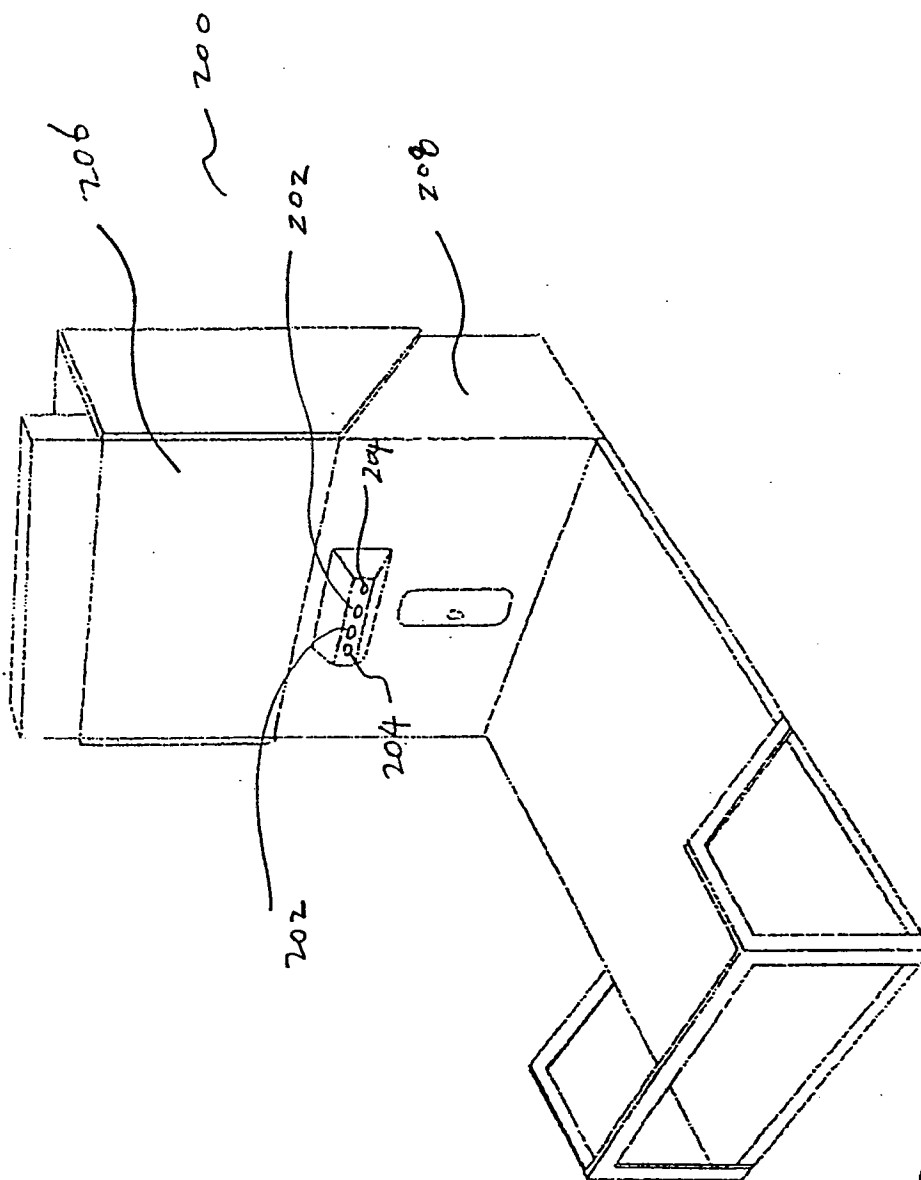


Fig. 8

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 99/27372

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 7 G06K9/20 G06T7/00

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HUBER E: "3-D real-time gesture recognition using proximity spaces" PROCEEDING. THIRD IEEE WORKSHOP ON APPLICATIONS OF COMPUTER VISION. WACV'96 (CAT. NO.96TB100084), PROCEEDINGS THIRD IEEE WORKSHOP ON APPLICATIONS OF COMPUTER VISION. WACV'96, SARASOTA, FL, USA, 2-4 DEC. 1996, pages 136-141, XP002135707 1996, Los Alamitos, CA, USA, IEEE Comput. Soc. Press, USA ISBN: 0-8186-7620-5 see paragraph 2.3  --- -/--	1-16

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

\* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

Date of mailing of the international search report

13 April 2000

28/04/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Sonius, M

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 99/27372

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>DARRELL T ET AL: "A virtual mirror interface using real-time robust face tracking"</p> <p>PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, XX, XX, 14 April 1998 (1998-04-14), pages 616-621, XP002083775</p> <p>see paragraph 2.1 -----</p>	1-16

CORRECTED VERSION

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
25 May 2000 (25.05.2000)

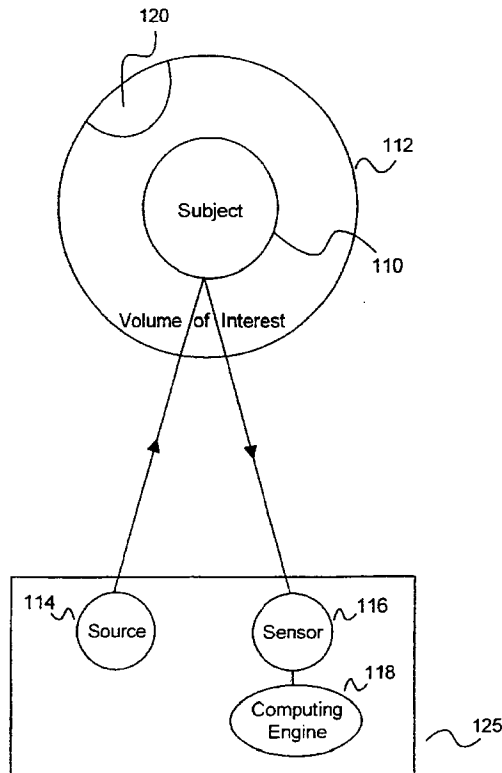
PCT

(10) International Publication Number  
**WO 00/030023 A1**

- (51) International Patent Classification<sup>7</sup>: **G06K 9/20**, (72) Inventors: **PU, Allen**; 1114 Las Riendas Way, Pasadena, CA 91107 (US). **QIAO, Yong**; 808 Wigwam Avenue, Arcadia, CA 91006 (US). **ESCOBAR, Nelson**; Apartment 302, 1310 E. Orange Grove Boulevard, Pasadena, CA 91104 (US). **ICHIRIU, Michael**; 472 S. El Molino Avenue #101, Pasadena, CA 91101 (US).
- (21) International Application Number: PCT/US99/27372
- (22) International Filing Date:  
17 November 1999 (17.11.1999)
- (25) Filing Language: English (74) Agents: **MEYER, Sheldon, R.** et al.; Fliesler Dubb Meyer and Lovejoy LLP, Suite 400, Four Embarcadero Center, San Francisco, CA 94111-4156 (US).
- (26) Publication Language: English
- (30) Priority Data:  
09/195,361 17 November 1998 (17.11.1998) US  
09/264,921 9 March 1999 (09.03.1999) US
- (81) Designated States (*national*): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT,
- (71) Applicant: **HOLOPLEX, INC.** [US/US]; Suite 102, 600 South Lake Avenue, Pasadena, CA 91106 (US).

[Continued on next page]

(54) Title: STEREO-VISION FOR GESTURE RECOGNITION



(57) Abstract: A method and an apparatus to identify a gesture of a subject without the need of a fixed background. The apparatus includes a sensor and a computing engine. The sensor captures images of the subject. The computing engine analyzes the captured images to determine 3-D profiles of the subject, and the gestures of the subject. Information in the images not within a volume of interest is ignored in identifying the gesture of the subject.

WO 00/030023 A1



RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA,  
UG, UZ, VN, YU, ZA, ZW.

(48) Date of publication of this corrected version:

29 August 2002

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(15) Information about Correction:

see PCT Gazette No. 35/2002 of 29 August 2002, Section II

**Published:**

--- with international search report

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

-1-

## STEREO-VISION FOR GESTURE RECOGNITION

BACKGROUND OF THE INVENTION

The present invention relates generally to gesture recognition, and more specifically to using stereo-vision for gesture recognition.

5 To identify the gestures of a subject, typically, the subject's background should be removed. One way to remove the background is to erect a wall behind the subject. After images of the subject are captured, the fixed background--the wall--is removed from the images before the gestures are identified.

It should be apparent from the foregoing that the wall increases the cost of the setup and the complexity to identify the gestures of the subject.

**SUBSTITUTE SHEET (RULE 26)**

-2-

SUMMARY OF THE INVENTION

The present invention identifies gestures of a subject without the need of a fixed background. One embodiment is through stereo-vision with a sensor capturing the images of the subject. Based on the images, and through ignoring  
5 information outside a volume of interest, a computing engine analyzes the images to construct 3-D profiles of the subject, and then identifies the gestures of the subject through the profiles. The volume of interest may be pre-defined, or may be defined by identifying a location related to the subject.

Only one sensor may be required. The sensor can capture the images  
10 through scanning, with the position of the sensor changed to capture each of the images. In analyzing the images, the positions of the sensor in capturing the images are taken into account. In another approach, the subject is illuminated by a source that generates a specific pattern. The images are then analyzed considering the amount of distortion in the pattern caused by the subject.

15 In one embodiment, the images are captured simultaneously by more than one sensor, with the position of at least one sensor relative to one other sensor being known.

In another embodiment, the subject has at least one foot, with the position of the at least one foot determined by a pressure-sensitive floor mat to help identify  
20 the subject's gesture.

The subject can be illuminated by infrared radiation, with the sensor being an infrared detector. The sensor can include a filter that passes the radiation.

In one embodiment, the volume of interest includes at least one region of interest, which, with the subject, includes a plurality of pixels. In analyzing the  
25 images to identify the gesture of the subject, the computing engine calculates the number of pixels of the subject overlapping the pixels of the at least one region of interest.



-3-

In another embodiment, the position and size of the at least one region of interest depend on a dimension of the subject, or a location of the subject.

In yet another embodiment, the present gesture of the subject depends on its prior gesture.

5        Other aspects and advantages of the present invention will become apparent from the following detailed description, which, when taken in conjunction with the accompanying drawings, illustrates by way of example the principles of the invention.

-4-

**BRIEF DESCRIPTION OF THE DRAWINGS**

FIG. 1 shows one embodiment illustrating a set of steps to implement the present invention.

FIG. 2 illustrates one embodiment of an apparatus of the present invention capturing an image of a subject.

FIG. 3 shows different embodiments of the present invention in capturing the images of the subject.

FIGS. 4A-C show one embodiment of the present invention based on the distortion of the specific pattern of a source.

FIG. 5 illustrates one embodiment of the present invention based on infrared radiation.

FIG. 6 shows different embodiments of the present invention in analyzing the captured images.

FIG. 7 shows one embodiment of a pressure-sensitive mat for the present invention.

FIG. 8 shows another embodiment of an apparatus to implement the present invention.

Same numerals in FIGS. 1-8 are assigned to similar elements in all the figures. Embodiments of the invention are discussed below with reference to FIGS. 1-8. However, those skilled in the art will readily appreciate that the detailed description given herein with respect to these figures is for explanatory purposes as the invention extends beyond these limited embodiments.

-5-

DETAILED DESCRIPTION

In one embodiment, the present invention isolates a subject from a background without depending on erecting a known background behind the subject. A three dimensional (3-D) profile of the subject is generated with the subject's gesture identified. The embodiment ignores information not within a volume of interest, where the subject probably is moving inside.

FIG. 1 shows one approach 100 of using an apparatus 125 shown in FIG. 2 to identify the gestures of the subject 110. At least one sensor 116, such as a video camera, captures (step 102) a number of images of the subject for a computing engine 118 to analyze (step 104) so as to identify gestures of the subject.

In one embodiment, the computing engine 118 does not take into consideration information in the images outside a volume of interest 112. For example, information in the images too far to the sides or too high can be ignored, which means that certain information is removed as a function of distance away from the sensor. Based on the volume of interest 112, the subject is isolated from its background. The gesture of the subject can be identified without the need of a fixed background.

FIG. 3 shows different embodiments of the present invention in capturing the images of the subject. One embodiment depends on using more than one sensor (step 150) to capture images simultaneously. In this embodiment, the position of at least one sensor relative to one other sensor is known. The position includes the orientation of the sensors. For example, one position is pointing at a certain direction, and another position is pointing at another direction.

Based on the images captured, the computing engine, 118, using standard stereo-vision algorithms, analyzes the captured images to isolate and to generate a 3-D profile of the subject. This can be done, for example, by comparing the

-6-

disparity between the more than one image captured simultaneously, and can be similar to the human visual system. The stereo-vision algorithm can compute 3-D information, such as the depth, or a distance away from a sensor, or a location related to the subject. That location can be the center of the subject. Information  
5 in the images too far to the sides or too high from the location can be ignored, which means that certain information is removed as a function of distance away from the sensors. In this way, the depth information can help to set the volume of interest, with information outside the volume not considered in subsequent computation. Based on the volume of interest, the subject can be isolated from its  
10 background.

In another embodiment, only one sensor is necessary. In one approach, the sensor captures more than one image at more than one position (step 152). For example, the sensor is a radar or a lidar, which measures returns. The radar can capture more than one image of the subject through scanning. This can be through  
15 rotating or moving the radar to capture an image of the subject at each position of the radar. In this embodiment, to generate the 3-D profile of the subject, the computing engine 118 takes into consideration the position of sensor when it captures an image. Before the subject has substantially changed his gesture, the sensor would have changed its position and captured another image. Based on the  
20 images, the 3-D profile of the subject is constructed. The construction process should be obvious to those skilled in the art. In one embodiment, the process is similar to those used in the synthetic aperture radar fields.

In another embodiment, the image captured to generate the profile of the subject depends on illuminating the subject by a source 114 that generates a specific  
25 pattern (step 154). For example, the light source can project lines or a grid of points. In analyzing the images, the computing engine considers the distortion of the pattern by the subject.

-7-

FIGS. 4A-C show one embodiment of the present invention depending on the distortion of the specific pattern of a source. FIG. 4A shows a light pattern of parallel lines, with the same spacing between lines, generated by a light source. As the distance from the light source increases, the spacing also increases. FIG. 4B shows a ball as an example of a 3-D object. FIG. 4C shows an example of the sensor measurement of the light pattern projected onto the ball. The distance of points on the ball from the sensor can be determined by the spacing of the projected lines around that point. A point in the vicinity of a smaller spacing is a point closer to the sensor.

In another embodiment, to enhance the ability in isolating the subject from the unknown background, the source 114 illuminates (step 160 in FIG. 5) the subject with infrared radiation, and the sensor 116 is an infrared sensor. The sensor may also include a filter that passes the radiation. For example, the 3dB bandwidth of the filter covers all of the frequencies of the source. With the infrared sensor, the effect of background noises, such as sunlight, is significantly diminished, increasing the signal-to-noise ratio.

FIG. 6 shows different embodiments of the present invention in analyzing the captured images. In one embodiment, the volume of interest 112 is predefined, 170. In other words, independent of the images captured, the computing engine 118 always ignores information in the captured images outside the same volume of interest to construct a 3-D profile of the subject.

After constructing the profiles of the subject, the computing engine 118 can determine the subject's gestures through a number of image-recognition techniques. In one embodiment, the subject's gestures can be determined by the distance between a certain part of the body and the sensors. For example, if the sensors are in front of the subject, a punch would be a gesture from the upper part of the body. That gesture is closer to the sensors than the position of the center of the body.

-8-

Similarly, a kick would be a gesture from the lower part of the body that is closer to the sensors.

In one embodiment, the volume of interest 112 includes more than one region of interest, 120. Each region of interest occupies a specific 3-D volume of space. In one approach, the computing engine 118 determines the gesture of the subject based on the regions of interest occupied by the 3-D profile of the subject. Each region of interest can be for designating a gesture. For example, one region can be located in front of the right-hand side of the subject's upper body. A part of the 3-D profile of the subject occupying that region implies the gesture of a right punch by the subject.

One embodiment to determine whether a region of interest has been occupied by the subject is based on pixels. The subject and the regions of interest can be represented by pixels distributed three dimensionally. The computing engine 118 determines whether a region is occupied by calculating the number of pixels of the subject overlapping the pixels of a region of interest. When a significant number of pixels of a region is overlapped, such as more than 20%, that region is occupied. Overlapping can be calculated by counting or by dot products. In another embodiment, the gesture of the subject is identified through edge detection, 173. The edges of the 3-D profile of the subject are tracked. When an edge of the subject falls onto a region of interest, that region of interest is occupied. Edge detection techniques should be obvious to those skilled in the art, and will not be further described in this application.

One embodiment uses information on at least one dimension of the subject, such as its height or size, to determine the position and size of at least one region of interest. For example, a child's arms and legs are typically shorter than an adult's. The regions of interest for punches and kicks should be smaller and closer to a child's body than to an adult's body. By scaling the regions of interest and by

-9-

setting the position of the regions of interest, based on, for example, the height of the subject, this embodiment is able to more accurately recognize the gestures of the subject.

5        This technique of modifying the regions of interest based on at least one dimension of the subject is not limited to three dimensional imaging. The technique can be applied, for example, to identify the gestures of a subject in two dimensional images. The idea is that after the 2-D profile of a subject is found from the captured images, the positions and sizes of two dimensional regions of interest can be modified based on, for example, the height of the profile.

10       Another embodiment sets the location of at least one region of interest based on tracking the position, such as the center, of the subject. This embodiment can more accurately identify the subject's gesture while the subject is moving. For example, when the computing engine has detected that the subject has moved to a forward position, the computing engine will move the region of interest for the kick  
15       gesture in the same direction. This, for example, reduces the possibility of identifying incorrectly a kick gesture when the body of the subject, rather than a foot or a leg, falls into the region of interest for the kick gesture. Identification of the movement of the subject can be through identifying the position of the center of the subject.

20       This technique of tracking the position of the subject to improve the accuracy in gesture recognition is also not limited to three dimensional imaging. The technique can be applied, for example, to identify the gestures of a subject in two dimensional images. The idea again is that after the 2-D profile of a subject is found from the captured images, the positions of two dimensional regions of interest  
25       can be modified based on, for example, the center of the profile.

In yet another embodiment, the computing engine takes into consideration a prior gesture of the subject to determine its present gesture. Remembering the

-10-

temporal characteristics of the gestures can improve the accuracy of gesture recognition. For example, a punch gesture may be detected when a certain part of the subject is determined to be located in the region of interest for a punch. However, if the subject kicks really high, the subject's leg might get into the region of interest for a punch. The computing engine may identify the gesture of a punch incorrectly. Such confusion may be alleviated if the computing engine also considers the temporal characteristics of gestures. For example, a gesture is identified as a punch only if the upper part of the subject extends into the region of interest for a punch. By tracking the prior position of body parts over a period of time, the computing engine enhances the accuracy of gesture recognition.

This technique of considering prior gestures to identify the current gesture of a subject again is not limited to 3-D imaging. For example, the technique is that after the 2-D profile of a subject is found from the captured images, the computing engine identifies the current gesture depending on the prior 2-D gesture of the subject.

FIG. 7 shows one embodiment of a pressure-sensitive floor mat 190 for the present invention. The floor mat further enhances the accuracy of identifying the subject's gesture based on the foot placement. In the above embodiments, the sensor 116 can identify the gestures. However, sometimes in situation, from the perspective of the sensor, when a certain part of the subject occludes another part of the subject, there might be false identification. For example, if there is only one sensing element, and the subject is standing directly in front of it, with one leg directly behind the other leg. Under this situation, it might be difficult for the computing engine to identify the gesture of the other leg stepping backwards. The pressure-sensitive mat 190 embedded in the floor of the embodiment 125 solves this potential problem.

In FIG. 7, the pressure sensitive mat is divided into nine areas, with a center



-11-

floor-region (Mat A) surrounded by eight peripheral floor-regions (Mat B) in four prime directions and the four diagonal directions. In this embodiment, the location of the foot does not have to be identified very precisely. When a floor-region is stepped on, a circuit is closed, providing an indication to the computing engine that a foot is in that region. In one embodiment, stepping on a specific floor-region can provide a signal to trigger a certain event.

In one embodiment, the volume of interest 112 is not predefined. The computing engine 118 analyzes the captured images to construct 3-D profiles of the subject and its environment. For example, the environment can include chairs and tables. Then, based on information regarding the characteristics of the profile of the subject, such as the subject should have an upright body with two arms and two legs, the computing engine identifies the subject from its environment. From the profile, the engine identifies a location related to the subject, such as the center of the subject. Based on the location, the computing engine defines the volume of interest, 172. Everything outside the volume is ignored in subsequent computation. For example, information regarding the chairs and tables will not be in subsequent computation.

FIG. 8 shows one embodiment 200 of an apparatus to implement the present invention. More than one infrared sensor 202 simultaneously capture images of the subject, illuminated by infrared sources 204. The infrared sensors have pre-installed infrared filters. After images have been captured, a computing engine 208 analyzes them to identify the subject's gestures. Different types of movements by the subject can be recognized, including body movement, such as jumping, crouching, leaning forward and backward; arm movements such as punching, climbing, and hand motions; and foot movements such as kicking, moving toward, and backward. Then, the gestures of the subject can be reproduced as the gestures of a video game figure shown on the screen of a monitor 206.

-12-

In one embodiment, the screen of the monitor 206 shown in FIG. 8 is 50 inches in diameter. Both sensors are of the same height from the ground and are four inches apart horizontally. In this embodiment, a pre-defined volume of interest is 4 feet wide, 7 feet long and 8 feet high, with the center of the volume being located at 3.5 feet away in front of the center of the sensors and 4 feet above the floor.

The present invention can be extended to identify the gestures of more than one subject. In one embodiment, there are two subjects, and they are spaced apart. Each has its own volume of interest, and the two volumes of interest do not intersect. The two subjects may play a game using an embodiment similar to the one shown in FIG. 8. As each subject moves, its gesture is recognized and reproduced as the gesture of a video game figure shown on the screen of the monitor 206. The two video game figures can interact in the game, controlled by the gestures of the subjects.

Techniques using, such as radar, lidar and cameras, have been described. Other techniques may be used to measure, such as depth information, which in turn can determine volume of interest. Such techniques include using an array of ultrasonic distance measurement devices, and an array of infrared LEDs or laser diodes and detectors.

Other embodiments of the invention will be apparent to those skilled in the art from a consideration of this specification or practice of the invention disclosed herein. It is intended that the specification and examples be considered as exemplary only, with the true scope and spirit of the invention being indicated by the following claims.

-13-

CLAIMSWe claim:

1. A method for obtaining information regarding a subject (110) without the need of a fixed background, the method comprising the steps of:  
capturing (102) images of the subject; and  
analyzing (104) the captured images without considering information in the images outside a volume of interest (112) for obtaining information regarding the subject.

2. A method as recited in claim 1 wherein:  
the method is for identifying at least one gesture of the subject (110);  
the step of analyzing is for identifying at least one gesture of the subject (110); and  
the images are captured simultaneously by more than one sensor (116), with the position of at least one sensor relative to one other sensor being known.

3. A method as claimed in any preceding claim wherein the volume of interest (112) is pre-defined.

4. A method as claimed in any preceding claims wherein:  
at least a part of the volume of interest (112) and at least a part of the subject (110) are represented by a plurality of pixels; and  
the step of analyzing includes the step of calculating the number of pixels of the subject (110) overlapping the pixels of the volume of interest (112) to obtain information regarding the subject (110).

-14-

1           5.     A method as recited in claims 1, 2 or 3, wherein the step of analyzing  
2 includes the steps of:  
3           identifying the profile of the subject (110) based on the images; and  
4           determining whether an edge of the profile of the subject (110) is within the  
5 volume of interest (112).

6  
7           6.     A method as recited in any preceding claims wherein the position of  
8 at least a part of the volume of interest (112) depends on at least one dimension of  
9 the subject (110).

10  
11          7.     A method as recited in any preceding claims wherein the size of at  
12 least a part of the volume of interest (112) is scaled based on at least one dimension  
13 of the subject (110).

14  
15          8.     A method as recited in any preceding claims wherein at least one  
16 position of the volume of interest (112) depends on one position of the subject  
17 (110).

18  
19          9.     An apparatus (125) for obtaining information regarding a subject  
20 (110) without the need of a fixed background, the apparatus (125) comprising:  
21           a sensor (116) configured to capture images of the subject; and  
22           a computing engine (118) configured to analyze the captured images  
23 without considering information in the images outside a volume of interest for  
24 obtaining information regarding the subject.

25  
26          10.    An apparatus (125) as recited in claim 9, wherein  
27 the apparatus is configured for identifying at least one gesture of the subject

-15-

1 (110);  
2 the computing engine is configured to analyze the captured image for  
3 identifying at least one gesture of the subject (110); and  
4 the apparatus further comprises at least one additional sensor to  
5 simultaneously capture images of the subject, with the position of at least one  
6 sensor relative to one other sensor being known.

7  
8 11. An apparatus (125) as recited in claims 9 or 10 wherein the volume  
9 of interest is pre-defined.

10  
11 12. An apparatus (125) as recited in claims 9, 10 or 11 wherein:  
12 at least a part of the volume of interest (112) and at least a part of the  
13 subject (110) are represented by a plurality of pixels; and  
14 the computing engine (125) is configured to calculate the number of pixels  
15 of the subject overlapping the pixels of the volume of interest (112) to obtain  
16 information regarding the subject (110).

17  
18 13. An apparatus (125) as recited in claims 9, 10 or 11 wherein the  
19 computing engine (125) is configured to  
20 identify the profile of the subject (110) based on the images; and  
21 determine whether an edge of the profile of the subject (110) is within the  
22 volume of interest (112).

23  
24 14. An apparatus (125) as recited in claims 9, 10, 11, 12 or 13 wherein  
25 the position of at least a part of the volume of interest (112) depends on at least one  
26 dimension of the subject (110).

27

-16-

1           15.    An apparatus (125) as recited in claims 9, 10, 11, 12, 13 or 14  
2    wherein the size of at least a part of the volume of interest (112) is scaled based on  
3    at least one dimension of the subject (110).  
4

5           16.    An apparatus (125) as recited in claims 9, 10, 11, 12, 13, 14 or 15  
6    wherein at least one position of the volume of interest (112) depends on one  
7    position of the subject (110).

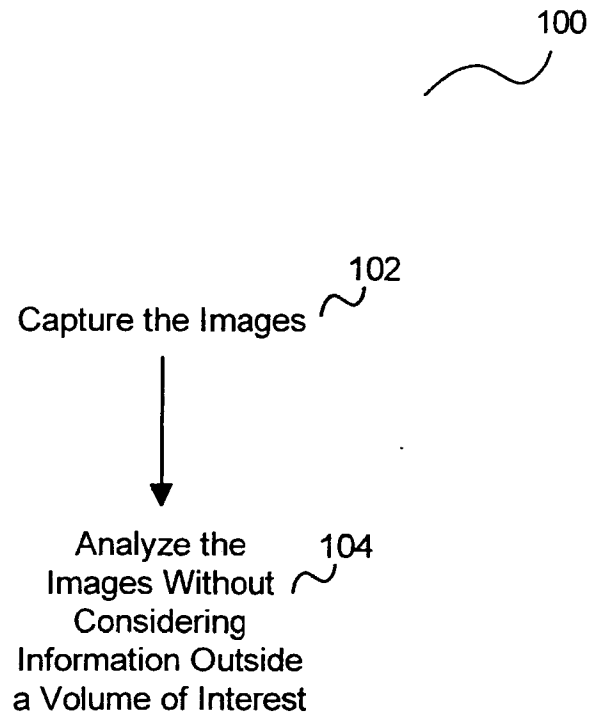


Figure 1

**SUBSTITUTE SHEET (RULE 26)**

2/8

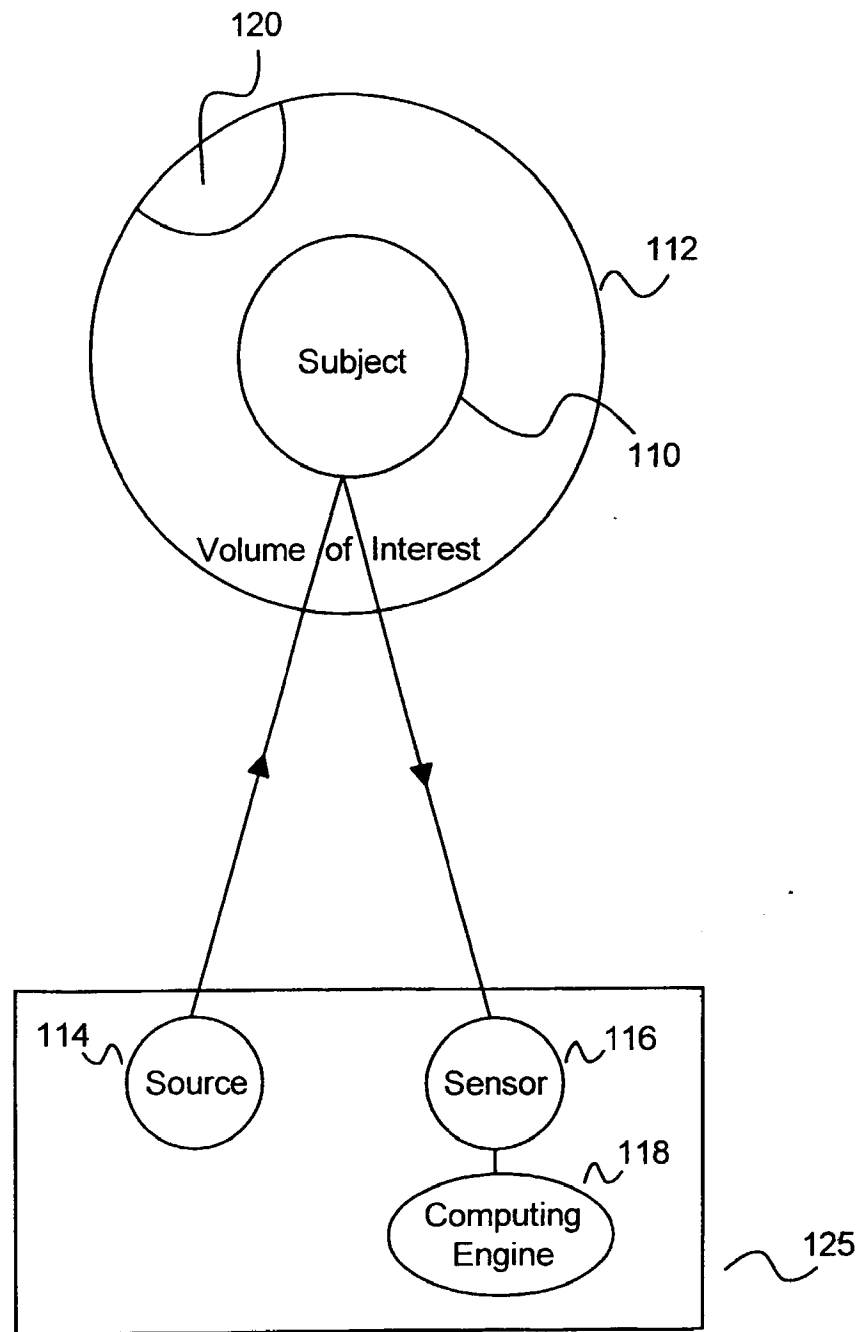


Figure 2

**SUBSTITUTE SHEET (RULE 26)**



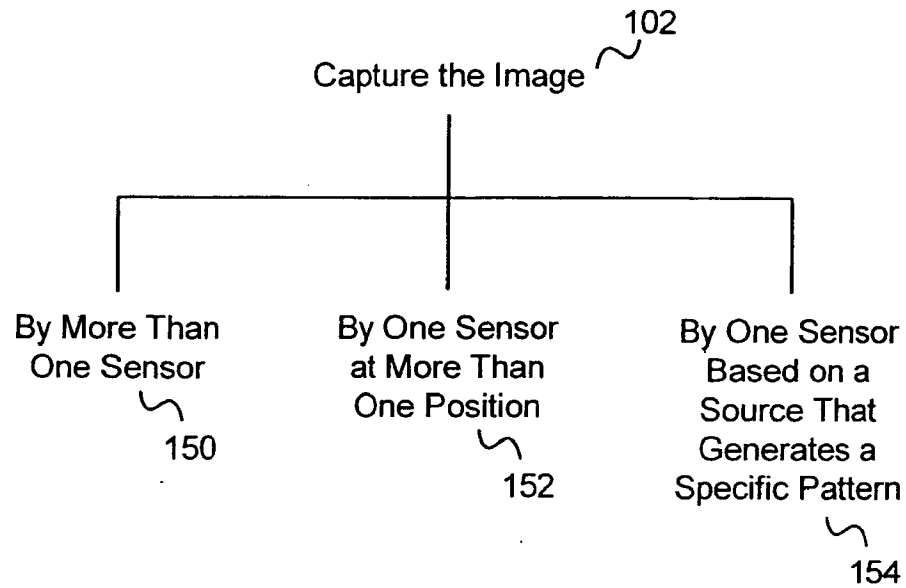


Figure 3

**SUBSTITUTE SHEET (RULE 26)**

4/8

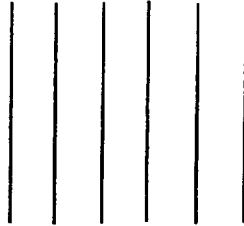


Figure 4A

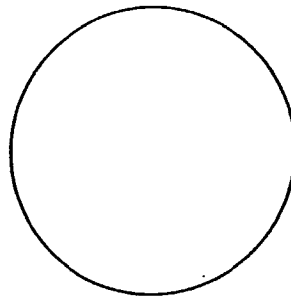


Figure 4B

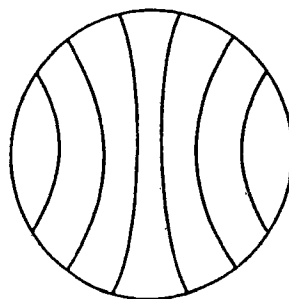


Figure 4C

**SUBSTITUTE SHEET (RULE 26)**

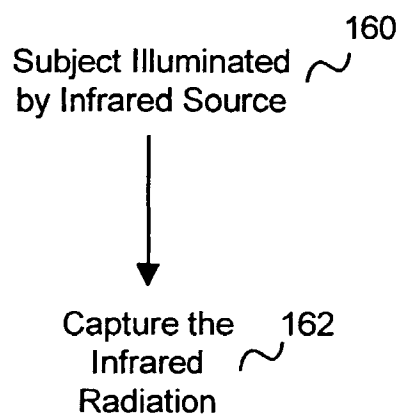


Figure 5

**SUBSTITUTE SHEET (RULE 26)**

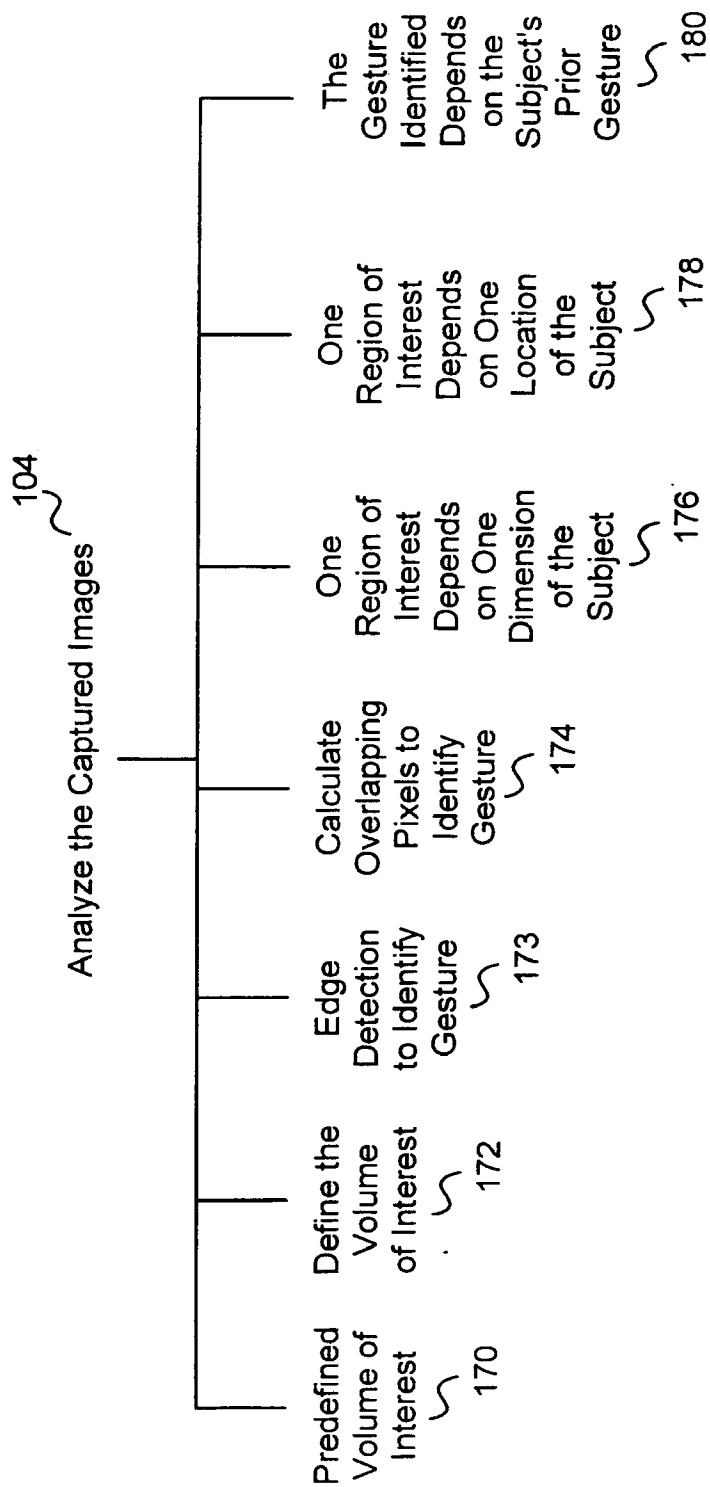


Figure 6

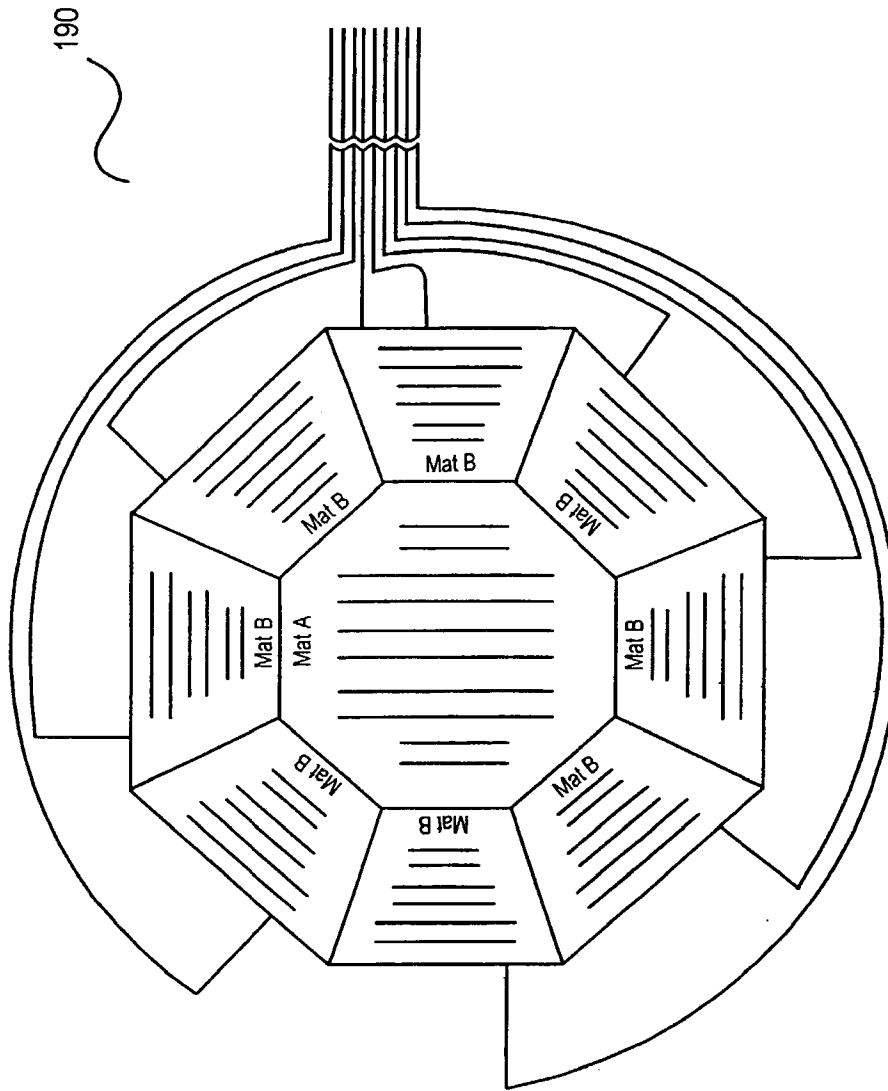


Figure 7

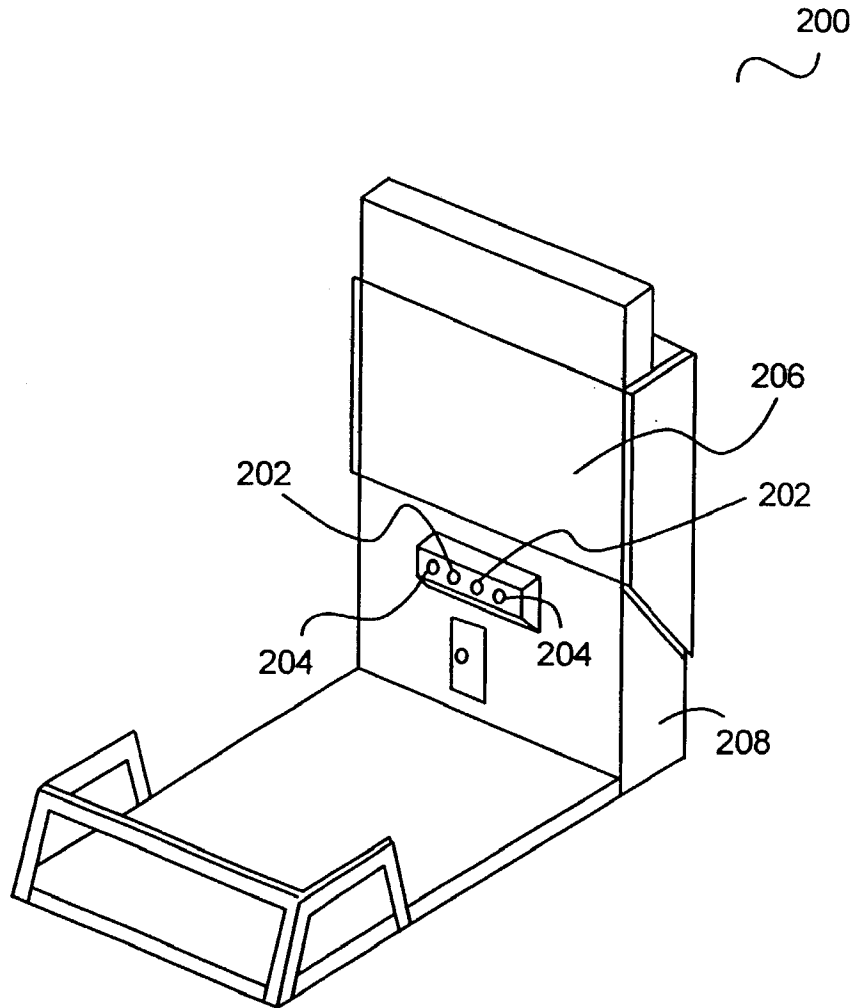


Figure 8

**SUBSTITUTE SHEET (RULE 26)**

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/27372

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 G06K9/20 G06T7/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HUBER E: "3-D real-time gesture recognition using proximity spaces" PROCEEDING. THIRD IEEE WORKSHOP ON APPLICATIONS OF COMPUTER VISION. WACV'96 (CAT. NO.96TB100084), PROCEEDINGS THIRD IEEE WORKSHOP ON APPLICATIONS OF COMPUTER VISION. WACV'96, SARASOTA, FL, USA, 2-4 DEC. 1996, pages 136-141, XP002135707 1996, Los Alamitos, CA, USA, IEEE Comput. Soc. Press, USA ISBN: 0-8186-7620-5 see paragraph 2.3 --- -/--	1-16



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

## \* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&amp;" document member of the same patent family

Date of the actual completion of the international search

13 April 2000

Date of mailing of the international search report

28/04/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Sonius, M

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 99/27372

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>DARRELL T ET AL: "A virtual mirror interface using real-time robust face tracking"</p> <p>PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, XX, XX, 14 April 1998 (1998-04-14), pages 616-621, XP002083775</p> <p>see paragraph 2.1</p> <p>-----</p>	1-16

Form PCT/ISA/210 (continuation of second sheet) (July 1992)